Complete Nucleotide Sequence of a Replicating RNA Molecule

The sequence suggests how nucleic acids exhibit phenotypes for selection and can evolve to greater complexity.

D. R. Mills, F. R. Kramer, and S. Spiegelman

We report here the complete sequence of an RNA molecule capable of extracellular replication. Use was made of "fragment length mapping," a new strategem that permits the unambiguous ordering of sequence blocks by determining which of several alternatives is closest to the block being extended.

A number of unexpected implications for the rules of precellular evolution emerged from the present study and we may briefly note them. The primary sequence contains a surprising number of intrastrand antiparallel complements, a peculiarity generating the potentiality for extensive secondary and tertiary structures containing antiparallel stems and loops. This possibility would allow these molecules to go beyond their primary sequences and exploit the selective advantages of their two- and three-dimensional consequences; a distinction between genotype and phenotype could thus arise before the primary sequence was used for translational purposes.

If, as seems likely, selection operates on secondary and tertiary structures, one can begin to identify the forces that could drive these molecules to greater length and complexity, a necessary prelude to the invention of cells and their components. Further, we will see how complementary copying and antiparallelism could have served to guide the evolution of the replicating single strands toward greater structural complexity. With the appearance of the first primitive cell the structural "phenotype" of the gene would become irrelevant as a selective element, and one could afford to store translatable genetic information in the perfectly paired double helix we know today.

The possibilities for such studies came in 1965 when we isolated (*1*, *2*) a template-specific RNA replicase from *Escherichia coli* infected with Qβ, an RNA bacteriophage, and established that the enzyme preparation could mediate a virtually indefinite autocatalytic synthesis (*3*) of biologically competent and infectious RNA (*4*). We further showed (*5*) that when Qβ replicase is presented wtih either of two genetically distinct Qβ-RNA molecules, the RNA synthesized is identical to the initiating template. This proved that the RNA is the instructive agent in the synthesis, thereby satisfying the operational definition of a self-instructive duplicating entity.

These findings opened up several novel experimental pathways. Potentially, one of the most interesting was that biologists were provided for the first time with an opportunity to explore Darwinian selection with nucleic acid molecules replicating outside a living cell. This situation simulates certain aspects of precellular evolution, when environmental discrimination presumably operated directly on the replicating gene, rather than on its translated product. The comparative simplicity of the system and the accessi-

The authors are members of the Department of Human Genetics and Development and of the Institute of Cancer Research, College of Physicians and Surgeons, Columbia University, New York 10032.

bility of its known chemical components to manipulation permit the imposition of a variety of defined selection pressures during growth of the replicating molecules.

In the test tube, the Qβ-RNA molecules are liberated from many of the restrictions stemming from the requirements of a complete viral life cycle. Thus, since replicase is provided, and since the RNA molecules need not infect cells for replication, the sequences coding for coat proteins and replicase might be dispensable. It was of obvious interest to design a selection experiment to test whether this possibility was in fact realizable. To this end a serial transfer experiment was performed (*6*) in which the selective advantage depended upon rapid completion of synthesis. The time between successive transfers was decreased as the molecules adapted by increasing their growth rates. The outcome was what might have been expected: the smaller the polynucleotide chain, the shorter the time required for completion. Accordingly, in response to the selection pressures imposed, the molecules eliminated dispensable sequences while retaining those required by the replicase for recognition and replication. As the experiment progressed, the molecules became smaller in a series of apparently discontinuous steps. By the 74th serial transfer, they had reached the lower size limit of 550 nucleotides attainable by this device. In doing so, these molecules (V-1 RNA) had discarded some 88 percent of the genetic information originally contained in the Qβ strands. That these abridged RNA variants were in fact derived from Qβ was established by hybridization experiments (*7*).

Variant molecules of the V-1 class were then transferred under conditions conferring advantages on molecules that could institute rapid synthesis when the reaction was initiated with as little as a single strand. In this manner, we isolated V-2 RNA, a strain of molecules that could be cloned (*8*). By variation of the specific nature of the selective pressures, other variants were derived from V-2, which could replicate at severely limiting concentrations of one of the required ribosidetriphosphates or in the presence of such inhibitory agents as tubercidin triphosphate (*9*), ethidium bromide, or proflavin (*10*).

It was evident from even the limited set of successful examples, and the ease with which they were attained, that the number of identifiably different mutant molecules possessing prespecified phenotypes was virtually unlimited. The only restriction would appear to stem from the ingenuity of the experimenter in devising the appropriate selective conditions. This diversity of phenotypes becomes understandable if the nucleic acid molecules can have a variety of secondary and tertiary structures (see below).

As we have noted previously (*11*), these experiments provided a plausible solution to the following dilemma of precellular evolution: "What pressures could have forced nucleic acid molecules toward greater length and complexity, a necessary prelude to the invention of cells or subcellular components?" It is possible to imagine interactions of mutant molecules and a primitive catalyst that would be similar to those discussed here. Nucleotide sequence changes that increased the catalytic capacity or accuracy even slightly would exert powerful selective effects in these precellular stages of evolving genetic material. Further, if, as seems likely, these interactions occur via secondary and tertiary structures, pressure to evolve toward greater length and complexity could be considerable.

In addition to their evolutionary interest, abbreviated molecules that have eliminated segments unnecessary to the enzyme for replication provide ideal objects for a detailed chemical analysis of the recognition and replication processes.

Search for a Replicating Molecule That Can Be Easily Sequenced

It became evident that a truly profound exploitation of the insight inherent in this in vitro system, and one that would bring us closer to the kind of chemical understanding we desired, demanded that we start selection with an RNA molecule of completely known sequence. To attain this goal, we sought for a replicating molecule with the following properties: (i) It should replicate in a manner recognizably analogous to that of Qβ-RNA (*7, 12–14*). (ii) It should be possible to isolate at least one of the two complementary strands for independent sequencing. (iii) It should be smaller than the 550-nucleotide length of the variants we had thus far studied so that its absolute sequence, and those of its mutants, could be obtained without inordinate time and effort.

We have reported (*15*) the isolation of a replicating RNA molecule (MDV-1; that is, midivariant-1) that contains 218 nucleotides and possesses the other features justifying the effort required to obtain its absolute sequence.

We now present the complete sequences of the plus and minus strands of this variant. In doing so we do not attempt to describe the extensive data accumulated, which led to the final sequences (*16*). Only the key experiments are summarized with particular emphasis on the new sequencing technology which we have devised.

Sequencing of Uniformly Labeled RNA Molecules

The basic pattern for sequencing RNA molecules introduced by Holley *et al.* (*17*) uses material synthesized in vivo and proceeds according to the following scheme: (i) Degrade the polynucleotide into oligonucleotides with the aid of the two endonucleases, pancreatic ribonuclease A and ribonuclease T_1. (ii) Isolate each oligonucleotide, a chore made very much easier by the extraordinarily convenient two-dimensional electrophoresis procedure of Sanger *et al.* (*18*). (iii) Determine the composition, sequence, and multiplicity of each oligonucleotide. (iv) Link the oligonucleotides into extended RNA segments by locating unique sequence overlaps among the oligonucleotides generated by ribonuclease A and ribonuclease T_1 digestion. (v) Finally, assemble the extended RNA segments into the complete sequence with the aid of an overlapping set of larger fragments isolated from partial nuclease digestions of the original RNA molecule. Step (iii) is lengthy and laborious with uniformly labeled molecules. Further, the large fragment isolation required in the last step often poses technical difficulties of considerable magnitude.

Strategy of Sequencing RNA Molecules Synthesizable in vitro

If the molecule being sequenced can be synthesized extracellularly, the logic of the sequencing procedure can be modified to ease the labor and, to a large extent, avoid the difficulties inherent in dealing with uniformly labeled material. The determination of the composition, multiplicity, and sequence of a given oligonucleotide is readily ac-

complished by running four separate reactions, each containing one of the ribosidetriphosphates labeled with ^{32}P in the α position and a fifth reaction containing all four ribosidetriphosphates labeled to the same specific activity. The fifth reaction serves to locate all the oligonucleotides produced and at the same time yields their composition. The four individually labeled reactions provide nearest-neighbor information (*19*) from which the sequences of many of the oligonucleotides can be directly deduced (*20*).

Use of Nearest-Neighbor Information

Here advantage is taken of the fact that in a cleavage of 3′,5′-phosphodiester bonds, whether by enzyme or alkali, the phosphorus of the resulting 2′ or 3′ nucleotides derives from the 5′ α-phosphorus of their nearest 3′ neighbors. If this type of analysis is carried out on a particular oligonucleotide derived from each of the four individually labeled syntheses, the data obtained almost always lead to the deduction of its unique sequence (*20*).

To exemplify this methodology, consider the ribonuclease A fragment pppGpGpGpGpApCp/C (*21*). This is of course the 5′ terminus of the original molecule since the terminal G carries the triphosphate group, a fact established by the presence in subsequent alkaline or T_1 hydrolyzates of the tetraphosphate pppGp. The composition of the fragment was determined from a uniformly labeled synthesis by separating and counting the ^{32}P-labeled nucleotides produced by alkaline hydrolysis. The results indicated that the fragment contained four residues of G and one each of A and C. The sequence of the six residues was then established from the individually labeled syntheses as follows. Alkaline hydrolysis of the terminal fragment derived from the synthesis containing [α-^{32}P]GTP revealed that only G residues were labeled with ^{32}P. Consequently, no G occurred on the 3′ side of either A or C, and all four G residues must therefore adjoin each other at the 5′ end of the fragment. The reaction containing [α-^{32}P]ATP yielded a fragment that contained one G residue with ^{32}P, establishing that the A residue follows immediately on the 3′ side of the four G residues. Finally, the A was labeled with ^{32}P only in the synthesis containing [α-^{32}P]CTP, placing the C after the A, thus completing the sequence of the fragment. Confirmation of this can be obtained by digestion of this fragment with ribonuclease T_1, which should, and does, lead to the production of ApCp/C.

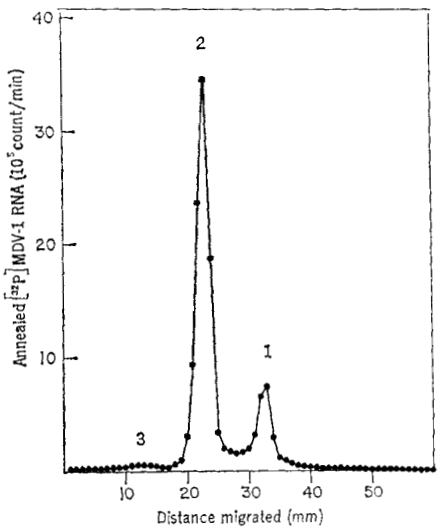

Fig. 1. Polyacrylamide gel electrophoresis of self-annealed MDV-1 RNA. The RNA product was synthesized and isolated (*15*). Prior to gel electrophoresis, the RNA was self-annealed for 60 minutes at 65°C in 400 m*M* NaCl, 3 m*M* ethylenediaminetetraacetic acid. The RNA species were then analyzed on 4.8 percent polyacrylamide gels (*24*) by electrophoresis at 10 ma per gel for 120 minutes. The RNA's from peaks 1 and 2 were eluted and precipitated with ethanol (*15*). Peak 1 contained only MDV-1 (+) RNA, the strand synthesized in excess. Peak 2 contained duplex structures formed by the annealing of the complementary MDV-1 (−) and MDV-1 (+) strands.

It should be further noted that nearest-neighbor information helps in ordering the oligonucleotides. For example, the 3′ terminal Cp of the 5′ terminal fragment was labeled with ^{32}P only in the synthesis containing [α-^{32}P]CTP. This implies that the next oligonucleotide on the 3′ side of this fragment must contain a C on its 5′ end. This type of information usefully limits the number of ways sequences can be assembled.

Use of Complementary Strands for Sequence Information

Another advantage derivable from an in vitro synthesis is the availability of the plus and minus strands produced in the synthetic reaction. We have shown that MDV-1 (*15*), like Qβ (*14*) and V-2 (*7*), replicates via a complementary antiparallel intermediate. Therefore, every purine-rich segment of the plus strand must be represented by an antiparallel pyrimidine-rich segment in the minus strand. By comparing oligonucleotides obtained from the plus strand with those from its complement, one can find overlapping complementary matches of sufficient length to extend and confirm known sequences.

Use of Synchronized in vitro Reactions

In their elegant sequence studies of Qβ-RNA, Weissmann and his colleagues (*22*) introduced the use of another benefit inherent in an in vitro synthesis. Knowing that the first six residues at the 5′ end of Qβ consist only of G and A, they initiated the synthesis at 37°C with a reaction mixture lacking CTP and UTP. This allows all the molecules to begin and then halt after the insertion of the first six residues. The temperature was then lowered to 20°C to discourage new initiations and to slow down subsequent polymerization. The missing CTP and UTP were then added, resulting in a synchronized onset of synthesis from the same point in all the molecules. Samples were taken at 5-second intervals, and the time required for a particular oligonucleotide to appear in the growing molecule was used to locate its position in the sequence. This procedure permitted the ordering of the first 300 nucleotides, although larger fragments obtained by partial nuclease digestion were still required to sort out the sequence.

Method of "Fragment Length Mapping" of Oligonucleotides

Ingenious as it is, the use of timed samples of synchronized polymerizations depends for success on two assumptions that are not always, or indefinitely, satisfied. One assumption demands that the replicase molecules traverse the polynucleotide template at constant speed during the sampling period, and the second requires that the reinitiated polymerizations remain in synchrony. The latter demand can hardly be maintained indefinitely, and a distance of about 300 nucleotides appears to be the limit. Beyond this, the synchrony becomes sufficiently blurred to interfere seriously with the resolving power of the method. The requirement of constant speed is likely to be satisfied only along polynucleotide stretches lacking significant secondary structures in the form of base-paired helices. This was a complication

that was probably aggravated in our instance, since, as will be seen, a major portion of our molecule is probably involved in such duplex-containing structures. Another difficulty originated from the 70 percent G + C content of our molecules, resulting in the production of a great many small oligonucleotides by digestion with ribonuclease.

In any event, we found that timed samples of synchronized reactions contained partially synthesized fragments of a rather broad size class that could not be used to arrange the oligonucleotides into an internally consistent linear array. We therefore devised a method of mapping by fragment length, a procedure that does not depend on constant rates of polymerization, avoids the ambiguities generated by loss of synchrony, and provides a higher degree of resolution. Examples of the method are described below, and we need here only note its simple logic. In any nucleic acid synthesis mediated by a nuclease-free polymerase, every fragment will contain the 5′ terminus. To locate a specific oligonucleotide, one need only determine the minimum fragment length required to ensure its presence. Separation of partially synthesized RNA fragments into narrow size classes can be achieved by electrophoresis through polyacrylamide gels.

We now describe each of the sequencing steps dictated by the particular strategy adopted. Emphasis will be focused on the plus and minus strand oligonucleotide catalogs, the construction of extended sequence blocks by the use of complementary oligonucleotides, the logic used in deciding which blocks can be joined, and finally on the "fragment length mapping" of the extended sequence blocks with the aid of high-resolution gel electrophoresis. We conclude with the complete sequences of the two complementary strands and a discussion of their probable secondary structures.

Table 1. MDV-1 RNA plus strand catalogs. The RNA from peak 1 was analyzed by digestion with ribonuclease T_1 and ribonuclease A. The oligonucleotide fragments obtained from these digestions are listed. The actual fragments are contained within the vertical bars. The residue outside each bar indicates the identity of the nearest-neighbor nucleotide. The nearest neighbor on the 5′ side of ribonuclease A fragments can be either C or U. The number of times each fragment occurs in the RNA is indicated to the left of its sequence. The occurrence of many small fragments is due to the high G + C content (70 percent) of the RNA.

Ribonuclease T_1 digest			Ribonuclease A digest		
5 G|G|C	2 G|ACG|C	2 G|CUG|C	11 §|U|C	1 §|AC|C	3 §|GAC|C
4 G|G|A	2 G|ACG|A	1 G|CUAG|C	1 §|U|A	4 §|AC|G	3 §|GAC|G
13 G|G|G	1 G|AAG|A	1 G|UACG|G	2 §|U|G	2 §|GC|C	1 §|AGC|C
4 G|G|U	2 G|AAG|G	½ G|CCUCG|A	6 §|U|U	3 §|GC|A	2 §|AGC|G
5 G|CG|C	1 G|CCCG|C	½ G|CCUCG|U	13 §|C|C	7 §|GC|G	1 §|GGC|C
3 G|CG|A	1 G|ACCG|U	1 G|UCACG|G	3 §|C|A	5 §|GC|U	1 §|GGC|G
1 G|CG|G	1 G|CACG|A	1 G|UUCG|A	13 §|C|G	1 §|GU|A	1 §|GGGC|A
2 G|CG|U	1 G|AACCG|C	1 G|CUUCG|C	5 §|C|U	3½ §|GU|G	1 §|GGGC|U
1½ G|AG|A	1 G|CCACG|C	1 G|UUUCG|G	1 §|C-OH	1 §|GU|U	1 §|GAGU|C
3 G|AG|G	2 G|UG|C	1 G|CUUUCG|C	1 pppGGGGAC|C		1 §|AGGU|G
2 G|AG|U	3½ G|UG|A	1 pppG|G	½ §|GAAGAGGC|G		2 §|GAGGU|G
2 G|CAG|C	1 G|UG|G	1 G|UUCCCC-OH	½ §|GAGAAGAGGC|G		1 §|GGGAGU|U
1 G|ACCCCCG|A		1 G|CCCUUCG|C	1 §|GAAGGGGGGU|U		1 §|GAGAAC|C
1 G|ACCCCCCG|G		1 G|ACCUUCG|U	1 §|GGAAGGGGGGAC|G		
1 G|CACCUCG|U		1 G|CUCUCCAG|G			

Synthesis of the Plus and Minus Strands of MDV-1 RNA

Figure 1 shows a 4.8 percent polyacrylamide gel profile of a representative self-annealed product from a reaction initiated with MDV-1 (+) RNA. We have previously shown (*15*) that peak 1 contains mature single strands, peak 2 consists of double-stranded duplexes of antiparallel complements, and peak 3 comprises multistranded complexes with both complementary strands present in various proportions. Since the material was self-annealed prior to electrophoretic separation, peak 1 contains the single strand made in excess, which we have designated as the plus strand, in accordance with the situation that obtains in the in vitro synthesis of Qβ-RNA (*14*). All three peaks are easily eluted from the gels in high yield for further characterization.

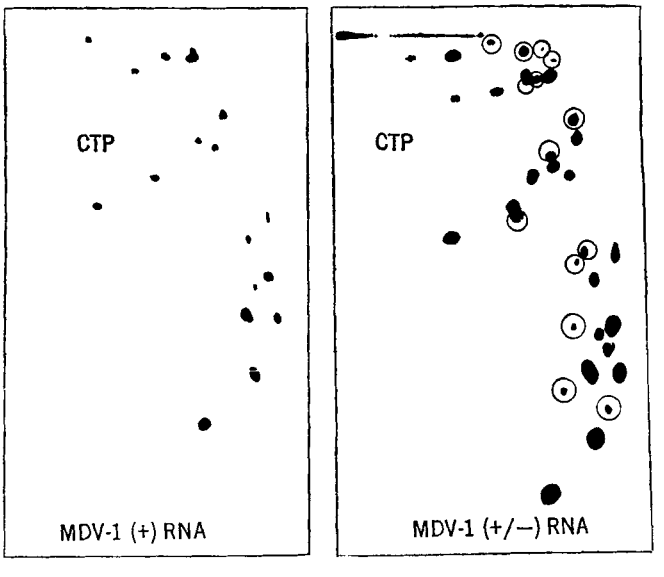

Fig. 2. Comparison of ribonuclease T_1 digests of MDV-1 (+) and MDV-1 (+/−) RNA. RNA from peak 2 was melted to separate the strands of the duplex into an equimolar mixture of plus and minus RNA, digested with ribonuclease, and analyzed by two-dimensional electrophoresis (*18*). The resulting two-dimensional nucleotide patterns (fingerprints) were more complicated than their plus strand counterparts, as they contained the fragments of both strands. The minus strand oligonucleotides were identified by a comparison of the duplex and single-stranded fingerprints. The ribonuclease T_1 fingerprints shown were obtained from [α-^{32}P]CTP-labeled RNA. Some of the fragments found only in the minus strand are circled. Many other minus strand oligonucleotides, however, are identical to those found in the plus strand.

Oligonucleotide Catalogs of MDV-1 (+) RNA

Four separate syntheses were run, each containing one of the four ribosidetriphosphates labeled with ^{32}P in the α position, and the RNA in peak 1 was isolated. Portions were subjected individually to either ribonuclease A or T_1 digestion. The oligonucleotides were then separated by the Sanger (*18*) two-dimensional electrophoresis procedure. In those few cases where oligonucleotides were not adequately resolved (that is, the very large oligonucleotides derived from ribonuclease A digestion), the digests were first separated according to size (isopleths) by urea-diethylaminoethyl cellulose chromatography (*23*), and then each isopleth was resolved into its component oligonucleotides by two-dimensional electrophoresis.

Every oligonucleotide from the four syntheses was eluted and subjected to alkaline hydrolysis. The resulting 2′ and 3′ nucleotides were separated electrophoretically, and the distribution of ^{32}P was determined, in order to yield the nearest-neighbor analysis of each residue.

As a further aid, each oligonucleotide was digested with the nuclease not used to produce it. Thus, T_1 oligonucleotides were digested with ribonuclease A and vice versa. Electrophoretic separation of these digests yielded predictable sequence elements of each oligonucleotide. They served to confirm, and in many cases helped to establish, the sequence of each oligonucleotide. In certain instances partial digestion with spleen phosphodiesterase (*18*) was also used to obtain sequence elements of oligonucleotides.

When the chain length of a particular oligonucleotide was in doubt, it was first exposed to alkaline phosphatase to remove charges due to terminal phosphates, and then chromatographed again on diethylaminoethyl cellulose with known chain length markers as described (*20*).

The 3′ terminal oligonucleotide was identified by alkaline hydrolysis of a product synthesized with [^{3}H]CTP. The 3′ oligonucleotide terminus yielded the nucleoside [^{3}H]cytidine, whereas the other fragments gave only nucleotides on hydrolysis. The 5′ terminus was identified as the only ribonuclease A fragment that was labeled by [γ-^{32}P]-GTP and by the fact that it yielded pppGp on alkaline hydrolysis.

Once the sequence of each oligonucleotide was ascertained, its molar frequency in the molecule was determined by comparing its ^{32}P content with that of the entire digest.

Table 1 summarizes the complete T_1 and ribonuclease A oligonucleotide catalogs determined in this manner. All oligonucleotides in this and subsequent tables and figures are written in the 5′ to 3′ direction. The residues outside the vertical bar on the 5′ side are deduced from the known properties of the enzyme used to obtain the fragment. The residues on the 3′ side of the vertical bar are derived from nearest-neighbor information. The multiplicity of each oligonucleotide is indicated to the left of its sequence. Of interest is the appearance of a few fragments with nonintegral values, a puzzling feature that was ultimately resolved by the discovery that our variant was a mixture of two mutants that differed at residue 104 of the plus strand.

Oligonucleotide Catalogs of MDV-1 (−) RNA

Since the complementary MDV-1 (−) RNA is not obtainable in adequate yield, the oligonucleotide catalogs of this species were obtained by a rather different device. The double-helical RNA (peak 2) of Fig. 1, containing both plus and minus strands, was eluted from the gel, concentrated, and melted. Portions of the denatured equimolar mixture of plus and minus strands [MDV-1 (+/−) RNA] were then digested individually with ribonuclease A and ribonuclease T_1 as described for MDV-1 (+) RNA. The resulting digest contained a mixture of fragments derived from plus and minus strands.

For example, in Fig. 2, the single-stranded (+) RNA and the double-stranded (+/−) RNA were labeled with [α-^{32}P]CTP and then digested with ribonuclease T_1. From these digests, two-dimensional electrophoretic patterns (fingerprints) were obtained. Some of the sequences found only in the minus strand are circled. Others are clearly common to both.

To sort out the set of oligonucleotides complementary to those found for the plus strand, four duplex preparations were synthesized, each labeled with a different riboside [α-^{32}P]triphosphate. The ribonuclease A and T_1 digests were then used to construct the complete oligonucleotide catalogs for the combined equimolar mixture of plus and minus strands. The MDV-1 (+) catalogs of Table 1 were then subtracted from the MDV-1 (+/−) catalogs to yield the ribonuclease A and T_1 catalogs of the minus strand listed in Table 2. Again, oligonucleotides with nonintegral molar frequencies are present, a consequence of the mixture of the two variants differing at one residue.

The plus and minus catalogs of Tables 1 and 2 were established virtually independently of each other. The complementarity of the plus and minus catalogs is evidence of their correctness.

Construction of Extended Sequence Blocks

The ribonuclease T_1 and A fragments of Tables 1 and 2 can be extended into larger sequence blocks by the following devices: (i) If a T_1 oligonucleotide uniquely overlaps a ribonuclease A oligonucleotide, the one may be used to extend the other. (ii) Nearest-neighbor residues on the 5′ end of oligonucleotides are deducible from the enzyme used, and nearest neighbors on the 3′ end are determined by identifying the riboside [α-^{32}P]triphosphate that labels the 3′ terminal phosphate. This nearest-neighbor information logically limits the choice of the fragments to be placed on either side of a given oligonucleotide. (iii) Since MDV-1 (+) and MDV-1 (−) are antiparallel complements, one can use sequences unique in one strand to extend sequences in the other.

For example, the plus strand catalog (Table 1) contains the unique ribonuclease A oligonucleotide Py/GAGAAC/C. Searching the minus strand catalog for unique antiparallel complements, one finds G/UUCUCG/U among the T_1 fragments. This immediately allows us to extend the plus ribonuclease A fragment to include ACGAGAACC. This last sequence can be further lengthened on the 3′ side by recognizing that a T_1 cleavage of this segment should yield a fragment beginning with G/AACC . . . /, and there is only one such T_1 fragment in the plus catalog, namely, G/AACCG/C. Since this must be on the 3′ side of ACGAGAACC, the sequence becomes ACGAGAACCGC.

By means of these devices, all of the oligonucleotides of the plus and minus catalogs were assembled into the extended sequence blocks recorded in Fig. 3. For convenient referencing the blocks are not arranged in the order of their identification in the course of

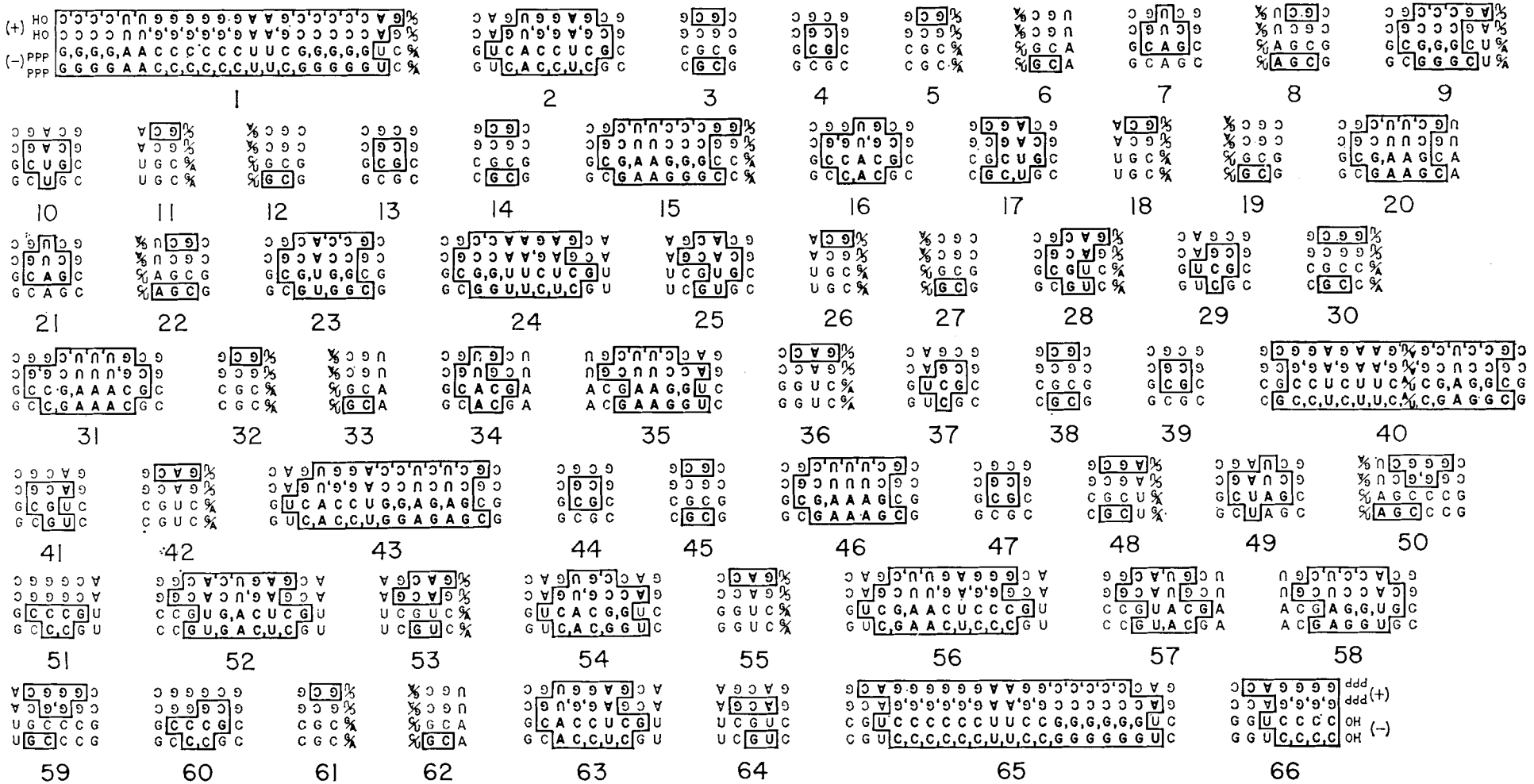

Fig. 3. The extended sequence blocks. Information obtained from enzymatic digestion (Tables 1 and 2) was combined to yield 66 sections of duplex MDV-1 RNA. Each strand is represented twice to accommodate fragments from the two enzymatic digests. The actual fragments are contained within the outline of each block, and the points of enzymatic cleavage are indicated by commas. In assembling the blocks, the smaller fragments of one catalog were "used up" by the larger fragments of another catalog. In some blocks, however, the overlapping of larger fragments (by complementarity or identity) extended the length of the block. The size of each block was limited by the proviso that only unique sequences could be used to extend it. If two or more different oligonucleotides could be added in the same location, the choice could not be made. The advantage of assembling catalog data into block form is that each block cannot be combined with every other block. The identity of each block's nearest neighbor is logically restricted to only a few other blocks. For instance, the right sides of blocks 24, 52, and 63 are identical. These blocks can only be linked with the left sides of blocks 25, 53, and 64. The problem at this point was to determine which blocks on the left paired with which blocks on the right. In all, there were 15 "logical choice sets," such as this one.

our analysis, but rather, according to their ultimate linear arrangement in the molecule.

The conventions used in representing these blocks may be briefly noted. Both minus (lower two) and plus (upper two) strands are represented, the latter being inverted to indicate the antiparallel structure of the two complements and to permit both to be read in the 5′ to 3′ direction. Plus and minus strands are represented twice so that the sequence information from the two nuclease catalogs could be explicitly indicated. The first and the fourth rows correspond to the ribonuclease A fragments, and the second and third rows correspond to the T_1 oligonucleotides, the commas indicating points of enzymatic cleavage. Only the residues enclosed within the outlined figures belong to the final sequence. The outside residues on either side of each block are deduced from identity, complementarity, enzyme specificity, and nearest-neighbor information. These external residues are brought in by the neighboring block. They, and the end shape of a block, help in choosing its nearest neighbor. The actual fragments that will "replace" the external residues of a block are to be found within the outlined figure of the adjoining block.

Table 2. MDV-1 RNA minus strand catalogs. These catalogs were prepared by deleting oligonucleotides found in each plus strand catalog from the corresponding catalog of fragments obtained from MDV-1 (+/−) RNA.

Ribonuclease T_1 digest			Ribonuclease A digest		
4 G|G|C	2 G|CAG|C	2 G|CUG|C	10½ §|U|C	3 §|AC|C	1 §|GAC|U
2 G|G|A	1 G|AAG|C	2 G|UCG|C	1 §|U|A	4½ §|AC|G	1 §|AGC|C
11 G|G|G	2 G|AAG|G	1 G|UCG|A	3 §|U|G	3 §|GC|C	2 §|AGC|G
6 G|G|U	1 G|CCCG|C	1 G|CUAG|C	3 §|U|U	3 §|GC|A	1 §|GGC|G
5 G|CG|C	1 G|CCCG|U	1 G|UACG|A	17 §|C|C	7 §|GC|G	1 §|GGU|C
3 G|CG|A	1 G|CACG|A	1 G|UCACG|G	4½ §|C|A	2 §|GC|U	1 §|GGU|U
2 G|CG|G	1 G|AAAG|C	1 G|ACUCG|U	13½ §|C|G	4 §|GU|C	1 §|GGGC|U
3 G|CG|U	1 G|CCACG|C	1 G|CACCUCG|U	8½ §|C|U	1 §|GU|A	1 §|GAAC|U
1 G|AG|C	1 G|AAACG|C	1 G|UCACCUG|G	1 §|C-OH	3 §|GU|G	1 §|GAGGC|G
1 G|AG|A	2 G|UG|C	1 G|UUCUCG|U	1 pppGGGGAAC|C		1 §|GAAGC|A
2 G|AG|G	1 G|UG|A	1 pppG|G	1 §|GAGGU|G		1 §|GAAAC|G
1 G|CCG|A	1 G|UG|G	1 G|UCCCC-OH	1 §|GGGGGU|C		1 §|GAAAGC|G
1 G|AACUCCCG|U	½ G|CCUCUUCUCG|A		1 §|GAAGGU|C		1 §|GAAGGGC|C
1 G|UCACCUCG|C	½ G|CCUCUUCACG|A		1 §|GGGGGGU|C		1 §|GGAGAGC|G
1 G|AACCCCCCUUCG|G	1 G|UCCCCCCUUCCG|G				

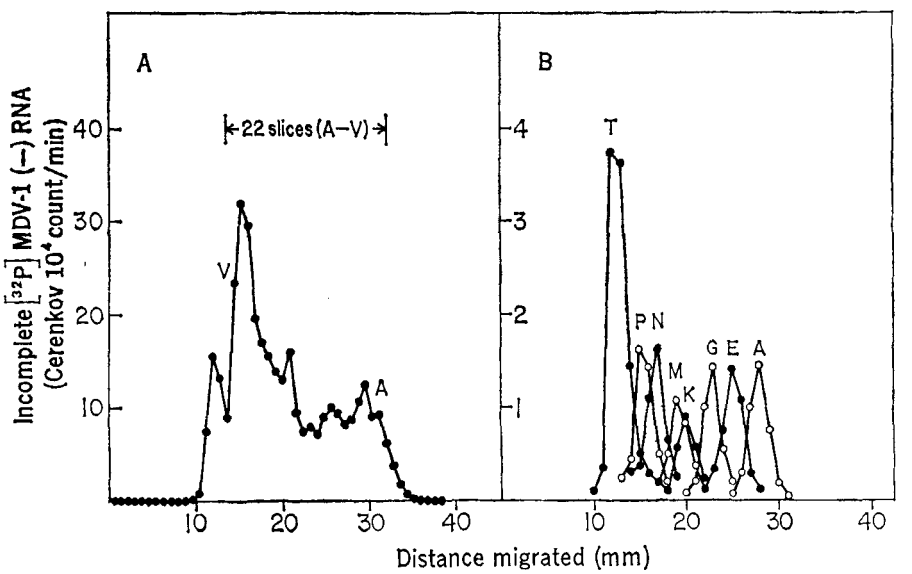

Fig. 4. Isolation of incomplete MDV-1 (−) RNA's by gel electrophoresis. (A) A reaction, with an excess of MDV-1 (+) RNA as template, was carried out at 38°C in the absence of CTP and UTP, to assure that chain initiation would occur without chain elongation beyond the initial GGGGAA sequence. The temperature was dropped to 20°C, and then CTP and UTP were added. Samples were taken at intervals as the chains elongated; the samples were pooled, resulting in a mixture of incomplete minus strands, all containing the 5′ end. The RNA was then purified, melted at 100°C, and separated according to length by electrophoresis through a 7.2 percent polyacrylamide gel. The RNA in each of 22 gel slices (labeled A through V) was isolated and subjected to electrophoresis again on 22 separate 7.2 percent gels. (B) The data from some of these second gels are plotted to illustrate the narrowness of the size classes isolated from each slice of the first gel. Moreover, only the RNA in the peak slice of each of the second gels was used for fragment length mapping. The resulting 22 RNA preparations were each exceptionally homogeneous in length.

Logical Restrictions in Connecting Sequence Blocks

When one strand of RNA is sequenced, nearest-neighbor information narrows the choice as to which oligonucleotide can be placed next to another. When double-stranded RNA is sequenced, nearest-neighbor data play a similar but far more selective role because each sequence block contains the two complementary strands and the demands made by the nearest neighbor of both strands must be satisfied. Two blocks can join together only if the external residues of the plus and minus strands of one block are precisely replaced by the identical internal residues of the plus and minus strands of the neighboring block.

For example, consider in Fig. 3 the problem of identifying the right-hand neighbor of block 34. Examination of all the sequence blocks reveals that only two blocks, 35 and 58, fulfill the requirements. Thus, the problem has been reduced to a choice between two alternatives. Once it is known which block pairs with block 34, the remaining block can be joined with the right side of block 57, as it is the only other block with a right side identical to block 34. Having narrowed the choice this far, it is only necessary to determine which pairs are close together in the molecule. To make this and similar choices, we developed the technique of "fragment length mapping."

"Fragment Length Mapping" of the Extended Sequence Blocks

We now describe the device for ordering sequence blocks that is independent of the time of their appearance or the rate of synthesis. Since our enzyme preparation is effectively free of

nuclease, every partially completed molecule in our reaction mixture will contain the 5′ terminus. If a collection of fragments synthesized haphazardly with respect to time or speed can be separated according to size, the following question can be answered: What is the minimum fragment length required to guarantee that it contain a specified oligonucleotide? If made sufficiently precise, it is evident that the answer locates the oligonucleotide in question in terms of distance from the 5′ end of the molecule. Separating RNA fragments according to size is achieved by electrophoresis through polyacrylamide gels (*24*). Further, the resolution capacities can be easily adjusted by varying gel strength and degree of cross-linking to suit the RNA size class.

To ensure a complete collection of fragment sizes, syntheses were carried out for varying periods, from 20 to 200 seconds. The products were then pooled and purified. To make certain that fragments of only the minus strand were synthesized, two precautions were taken. The reactions were carried out with a vast excess of the plus strand to minimize initiation on any completed minus strands. Further, after the 37°C initiation step, the temperature was dropped to 20°C to discourage any new starts. Fingerprint patterns of the pooled products revealed only minus strand oligonucleotides.

The mixture of products was subjected to electrophoresis through a 7.2 percent polyacrylamide gel and yielded the size classes shown in Fig. 4A. Each of the 22 gel slices was assigned a letter. The smallest (A) contained fragments of approximately 30 nucleotides in length and the largest (V) consisted of completed chains.

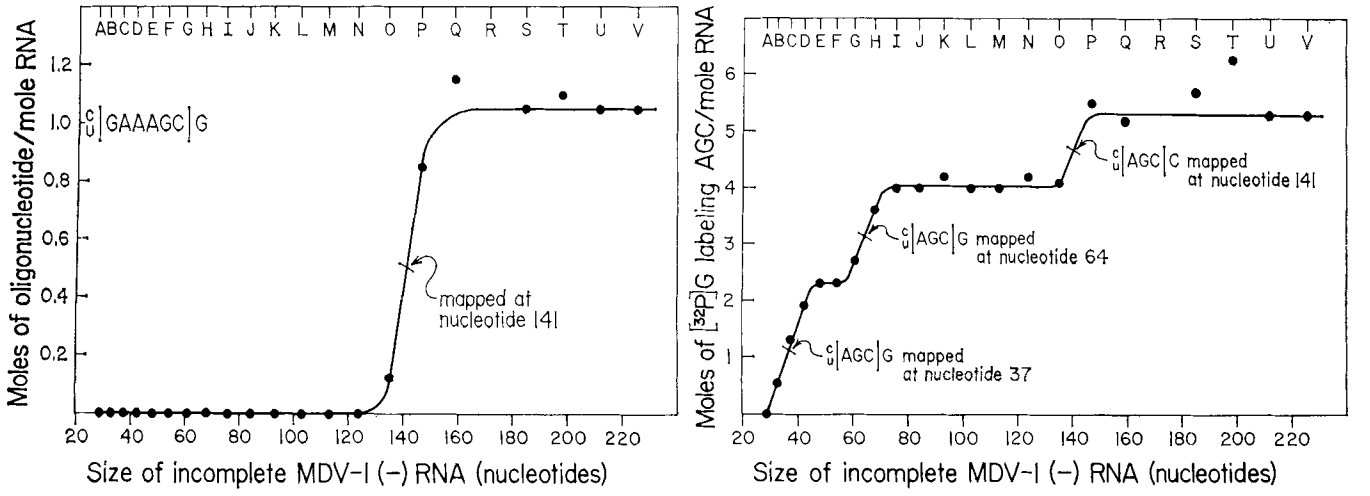

Fig. 5 (left). Fragment length mapping of an oligonucleotide. Each of the 22 size classes of incomplete MDV-1 (−) RNA was digested with nuclease and analyzed by two-dimensional electrophoresis. The sequence GAAAGC is found as a unique spot in the ribonuclease A fingerprint. This spot is not present in digests of incomplete RNA's smaller than size class N, but is fully present in digests of RNA's larger than size class P. By measuring the number of moles of GAAAGC present in each fingerprint, it was found that a minimum fragment length of 141 nucleotides was required to assure the presence of this oligonucleotide in the molecule. Fig. 6 (right). Fragment length mapping of three AGC oligonucleotides. The mapping technique easily resolved the three rises resulting from the occurrence of AGC three times in the minus strand. The first two AGC fragments to appear had a 3′ neighbor G and therefore each contained 2 moles of [^{32}P]G. Oligonucleotides were eluted from the AGC spot in each of the 22 fingerprint patterns and analyzed by alkaline hydrolysis. The results confirmed that the last of the three AGC residues to appear was the one with a C neighbor.

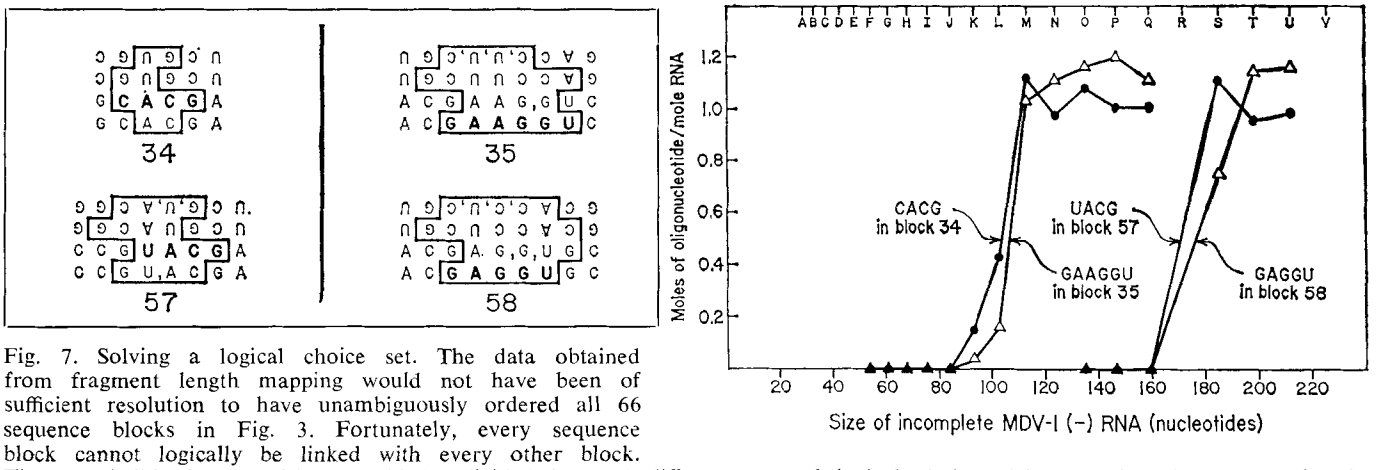

Fig. 7. Solving a logical choice set. The data obtained from fragment length mapping would not have been of sufficient resolution to have unambiguously ordered all 66 sequence blocks in Fig. 3. Fortunately, every sequence block cannot logically be linked with every other block. The permissible block pairings could be divided into 15 different sets of logical choices. The mapping data were of sufficient resolution to link the blocks in each of these choice sets. One such set is illustrated. Each block contains a unique oligonucleotide (indicated in bold type), whose position in the minus strand was determined by fragment length mapping. Block 34 could logically pair with either block 35 or block 58; however, its marker oligonucleotide CACG was mapped close to GAAGGU, the marker of block 35. In addition the markers of blocks 57 and 58 map close to each other, thus solving this logical choice set.

An interesting feature of the size distribution is the discontinuity evident in the accumulation of particular size classes. These irregularities can be explained by one or more of the following: (i) Aborted syntheses might occur with higher probability at certain sites in the template. (ii) Particular regions might be more difficult for the enzyme to traverse, thus slowing it down. (iii) The electrophoretic mobility of molecules in different states of completion may not be a smooth function of length. All three explanations invoke secondary structure. In any event, the observed bias toward certain size classes explains the difficulties noted earlier in ordering timed samples from synchronized syntheses. However, the bias toward certain size classes did not interfere with the use of fragment size to locate sequence blocks.

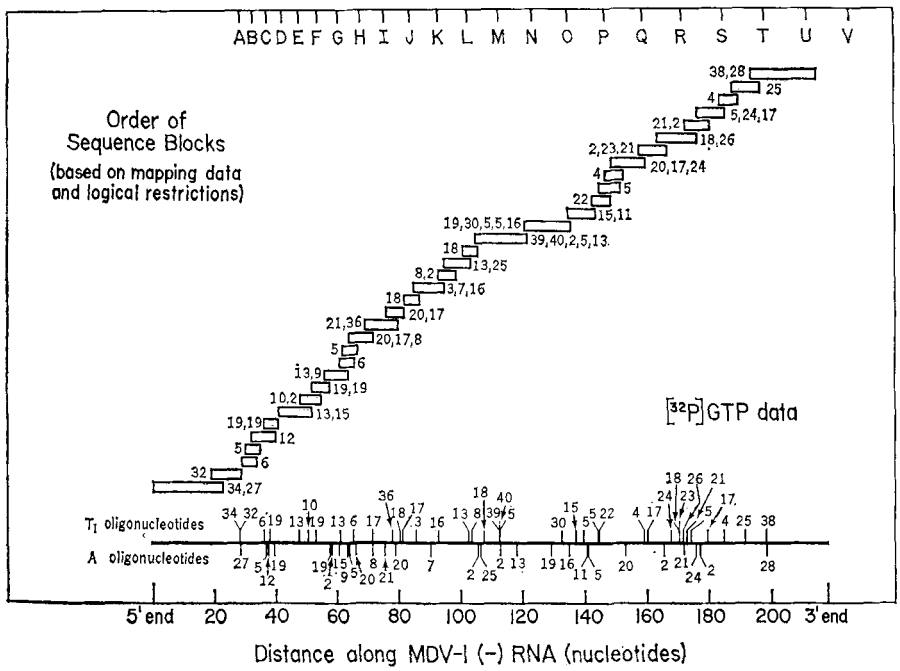

Fig. 8. Ordering the sequence blocks. The final position of the larger blocks in the sequence is indicated above. The key oligonucleotides used to map each block are indicated by the code numbers adjacent to each rectangle. The positions at which the key oligonucleotides appeared are indicated by the location of the codes on the map in the lower portion of the diagram. The mapping data were not of sufficient resolution to unambiguously locate the smaller blocks. However, the logical restrictions on block pairings, in combination with the known map positions of each block, permitted the final sequence to be assembled. Subsequent examination of fragments obtained by partial digestion of the molecule with ribonuclease, although not necessary, confirmed the final sequence.

The products contained in each of the 22 gel slices were eluted separately and precipitated with ethanol, and each was run again on another gel (Fig. 4B). The RNA's in the peak slices of each of these 22 gels were eluted and purified.

In this manner, quite homogeneous size classes of MDV-1 (−) fragments, all containing the 5′ terminus, were collected. Each of these 22 size classes was divided into two portions for digestion, one with ribonuclease T_1 and the other with ribonuclease A. The molar frequency of each oligonucleotide present in the resulting two-dimensional peptide patterns was determined for every size class.

We illustrate how this information was used for ordering sequence blocks with several examples. Consider the ribonuclease A oligonucleotide GAAAGC, which occurs only once in the minus strand (Table 2) and is a unique marker of sequence block 46 of Fig. 3. It is evident from Fig. 5 that GAAAGC rises from a molar frequency of zero to virtually one between the size classes of N and P. It is clear from the normalization to residue number (lower portion of Fig. 5) that this corresponds to an average chain length of 141. We can thus locate block 46 close to the 141st nucleotide of the minus strand.

A more complicated situation is illustrated with the oligonucleotide AGC, which occurs three times in the minus strand. The AGC sequences in blocks 8 and 22 have a 3′ neighbor G, while the AGC in block 50 has a C neighbor. Figure 6 shows that, as expected, the number of moles of [^{32}P]G that label

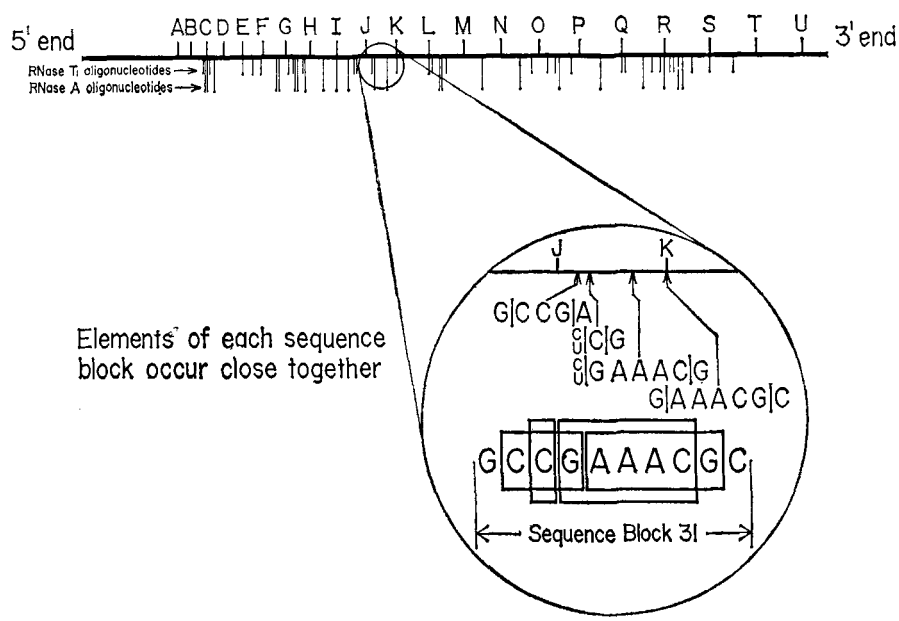

Fig. 9. High resolution within the map. Confirmation of the identity and location of each block. The oligonucleotides of each sequence block mapped close together. In no case were the predicted components of a block found in distant map locations. This confirmed the sequence of each block, since the oligonucleotides were assembled into blocks without any knowledge of their map locations. For example, the oligonucleotides of block 31, shown above, were all mapped between size classes J and K. Furthermore, the close mapping of many oligonucleotides from the same block increased the certainty with which a block could be located in the sequence.

AGC go through three rises. The two oligonucleotides with G neighbors appear first, at average chain lengths of 37 and 64, respectively. The last AGC, with a C neighbor, was mapped at nucleotide 141, which is in the second half of the molecule.

In discussing how sequence blocks are connected, we pointed out that nearest-neighbor information and shape of the ends impose logical restrictions that narrow the possibilities to comparatively few choices. In the particular example chosen, we found that either block 35 or block 58 could fit on the right side of block 34. By similar reasoning, block 35 or block 58 could fit on the right side of block 57. We now illustrate how the correct choice was made by the use of fragment length mapping.

In what follows, the oligonucleotides and the partially synthesized fragments being analyzed belong to the minus strand. Each of the four blocks we are concerned with contains a unique and identifying oligonucleotide. Thus, block 34 contains the ribonuclease T_1 fragment CACG, block 35 includes the ribonuclease A piece GAAGGU, block 37 is identified by the ribonuclease T_1 segment UACG, and finally block 58 contains GAGGU, a ribonuclease A oligonucleotide. Since each of these occurs only once in the minus strand catalog, each will undergo precisely one rise as a function of fragment length. Those that occur in neighboring blocks should make their appearance together in fragments of about equal minimum length. Figure 7 shows the minimum fragment length analyses for the identifying oligonucleotides and illustrates the certainty with which the correct choice can be made. It is clear that CACG is close to GAAGGU and that UACG is a neighbor of GAGGU. The results link block 34 to 35 and block 57 to 58.

There were 14 other sets of choices, many involving more than two possibilities, that had to be solved. However, in all cases the fragment length procedure yielded unique solutions. Figure 8 shows the map obtained from oligonucleotides labeled with [α-^{32}P]GTP. Additional experiments were also performed using the other three riboside [α-^{32}P]triphosphates. The resulting data, in conjunction with the logical restrictions imposed by the sequence blocks, yielded information that led to an unambiguous linear arrangement of the blocks, as shown in Fig. 8.

As an example of the high resolving power of the method, Fig. 9 shows an expanded region of the fragment length map. Sequence block 31 logically contained minus strand ribonuclease T_1 fragments CCG/A and AAACG/C, as well as ribonuclease A fragments C/G and GAAAC/G. The expanded map between size classes J and K shows the occurrence of each of the predicted elements. It must be emphasized that in this type of analysis the attempt is not to order the individual oligonucleotides, but rather the sequence blocks they identify. In no case, however, did the resulting map locations of the oligonucleotides suggest that the already established internal sequences of the blocks were not correct. Once the order of the sequence blocks had been determined, the final sequences were known. Figure 10 shows the complete nucleotide sequences of both strands of MDV-1 RNA.

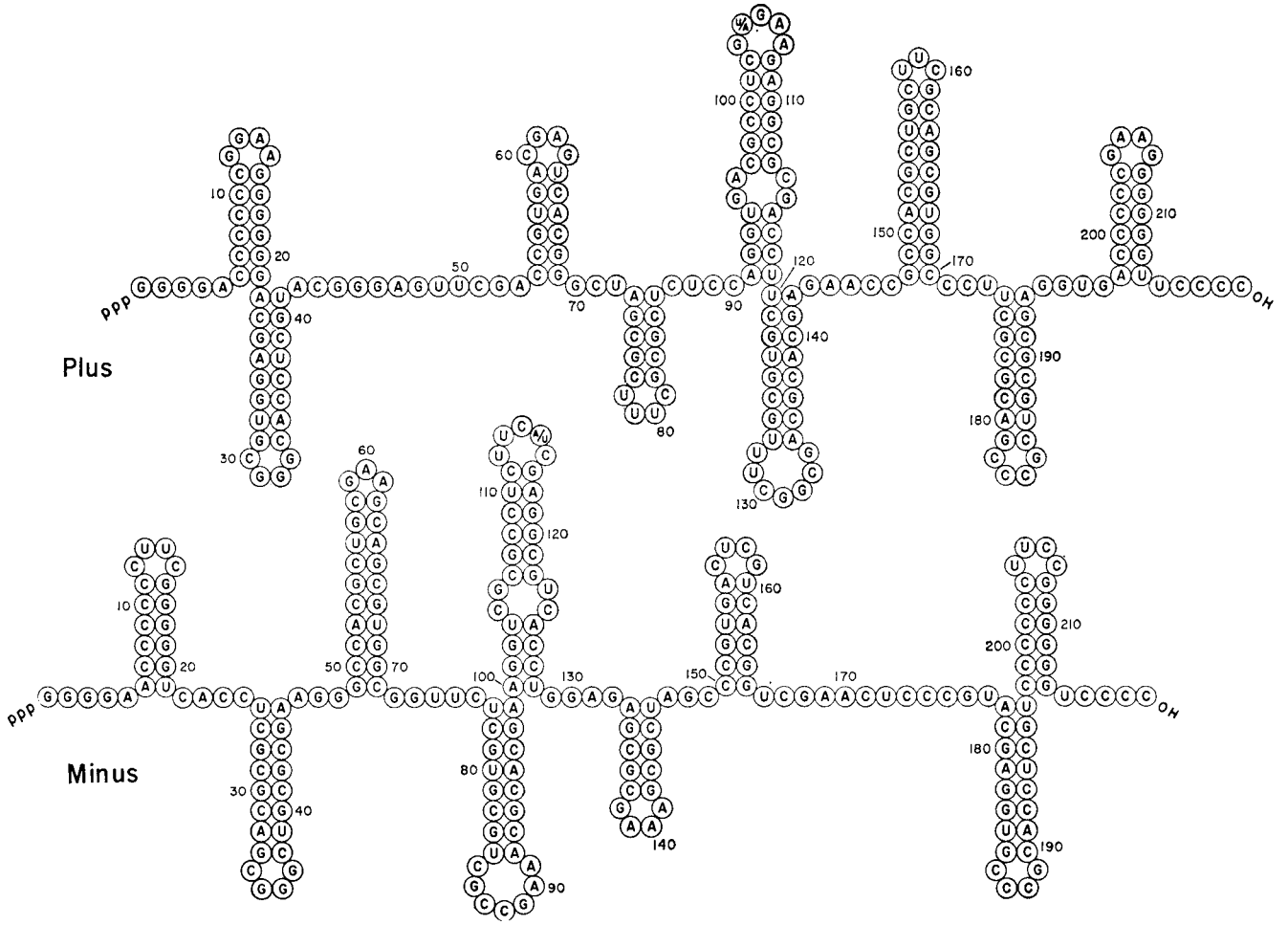

Fig. 10. The complete sequences of the complementary strands of MDV-1 RNA.

Existence and Implications of Secondary Structure

In the course of our studies a variety of features appeared pointing to the conclusion that the plus and minus strands were highly structured. Both complements had a high G + C content (70 percent) and a self-complementary nucleotide composition, with guanosine and adenosine each being equal in molar frequency to their respective complements. We also observed that the plus strand was very resistant (70 percent) to ribonuclease A under salt conditions that encouraged base pairing. Further, during our sequence studies we noted that some T_1 oligonucleotides were found in lower-than-expected yield at their correct position in the fingerprint, the missing portions remaining in the nuclease-resistant "core" at the top of the pattern. Finally, we discovered that a number of large oligonucleotides occurred in both the plus and minus catalogs. Any oligonucleotide common to both strands must of necessity imply the presence of its complement in both strands. This immediately generated the possibility of forming secondary structures involving pairing of the corresponding complementary regions in the same strand. Indeed, examination of the complete sequences made it likely that all of these expectations are in fact realizable. The oligonucleotides common to both strands were found to neighbor their intrastrand complements, increasing the possibility that they would complex. Furthermore, the deficient ribonuclease T_1 oligonucleotides were located in regions characterized by the possibility of forming hairpin structures containing loops lacking G residues and therefore resistant to ribonuclease T_1 digestion.

In summary, everything we knew of the detailed chemistry of the two strands dictated the probable existence of extensive base pairing, and this has been taken into account in the representations of the sequences shown in Fig. 10. It is important to emphasize that the plus and minus strand structures shown are based on comparatively simple rules and suggest only some of the many possibilities. Base pairing was assumed to occur only between neighboring complementary regions, separated by between three to seven residues; and providing the helical stem involves four or more complementary pairs.

More complicated structures than those of Fig. 10 can be readily produced (*25*) with the aid of bulge loops (*26*), G • U pairs (*27*), or stem formation involving distant complementary sequences. Further experiments are necessary to verify the reality of the loops shown and to see whether evidence for even more complex structures can be obtained.

Some Evolutionary Consequences of Complementarity

It is plausible to argue that the ability of complementary bases to pair by way of hydrogen bonds made possible the emergence of nucleic acids as self-instructive replicating entities. The use of base pairing demands that replication occur by way of complementary intermediates and this in turn introduces restraints on the evolutionary pathway. Thus, if a given variant is superior because of a particular sequence in its plus strand, optimization of the selective advantage would encourage the inclusion of the same sequence in the minus strand. This can be assured only if the plus strand also contains the antiparallel complement of the advantageous sequence. Note, however, that as soon as this device is employed both strands will necessarily contain intrastrand complements, and a new possibility emerges, namely, the formation of helices between the intrastrand antiparallel complementary regions. The recognition and use of such structures literally adds new dimensions to the types of selective interactions that can occur. The replicating molecules would no longer be limited to exploiting differences in the linear properties of their sequences but could in addition explore the usefulness of their two- and three-dimensional structure.

In the course of evolution, advantageous structures would be conserved. A structure retained in one strand would be expected to occur in the other. Thus, two complementary strands can have very different primary sequences (genotypes) and yet possess similar secondary and tertiary structures (phenotypes). As a result of this evolutionary process, self-replicating molecules would become highly structured and would possess many regions of intrastrand complementarity.

The double-stranded DNA we know today could not have participated in this kind of evolutionary game. A replicating molecule, such as DNA, probably arose later when environmental selection operated not on the gene but on the gene product. Under these circumstances secondary structures of the genetic material would become irrelevant.

The structures described in Fig. 10 illustrate many of the consequences of intrastrand complementarity. The helical portion of each hairpin in the plus strand has an identical counterpart in a complementary stem in the minus strand. For example, the helix between nucleotides 148 and 170 of the plus strand is identical to the helix in the minus strand between nucleotides 49 and 71. Further, it will be noted that the 5′ and 3′ ends of each strand are extensively complementary. One can therefore construct a distant intrastrand match involving 16 of the 20 nucleotides of each end. The fact that this, and other, structures can be formed with both strands is related to the ability of the replicase to accept either strand as a suitable template. It is of interest to recall that an end-to-end duplex of this nature was proposed to explain the observation (*2*) that fragments of Qβ-RNA could not be duplicated by Qβ replicase.

Although we have stressed the similarities between the complementary strands, we must also mention the differences. The two strands are actually mirror images of each other, in the sense that a helix at the 5′ end of one strand will generate an identical helix at the 3′ end of the other strand. Moreover, single-stranded sequences that do not have an intrastrand antiparallel complement cannot be represented in the other strand. Finally, if G • U pairs occur in one strand, the complementary C and A cannot form a pair in the other strand. The synthesis of more plus strand than minus strand (*13, 15, 20*) could be mediated by such differences between the strands.

The most striking characteristic of MDV-1 is its extensive secondary structure, a feature to be expected of a highly evolved replicating molecule. Indeed, if MDV-1 had been brought back in a moon sample, its sequence would have identified it not as a random polymer, but either as a self-instructive replicative entity or a product of one.

Concluding Comments

The immediate aim of the foregoing investigation was to determine the absolute sequence of a replicating nucleic acid molecule, and this has been achieved.

With the primary sequence known, one can hope to probe the chemistry of

the recognition device used by the replicase in selecting molecules for replication. In addition, the progress and details of the replication mechanism can be examined with far greater certainty of obtaining precise interpretable information. Finally, and most important, are its implications for extracellular Darwinian experiments. The fact that we can now start such selection experiments with a molecule of completely known sequence means that we can finally exploit the information inherent in this experimental system. For the first time, we can now ask, and answer, the following question: "Precisely what base changes have occurred in mutating from one phenotype to another?"

We have already shown (*28*) that, small as it is, MDV-1 RNA yields mutants with a prespecified phenotype, and one of these has been sequenced. It would be surprising if the data gained from such studies do not lead to new and perhaps unforeseen insights into the evolutionary pathways available to replicating nucleic acids.

References and Notes

1. I. Haruna and S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.A.* **54**, 579 (1965).
2. ———, *ibid.*, p. 1189.
3. ———, *Science* **150**, 884 (1965).
4. S. Spiegelman, I. Haruna, I. B. Holland, G. Beaudreau, D. Mills, *Proc. Nat. Acad. Sci. U.S.A.* **54**, 919 (1965).
5. N. R. Pace and S. Spiegelman, *Science* **153**, 64 (1966).
6. D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.A.* **58**, 217 (1967).
7. D. R. Mills, D. H. L. Bishop, S. Spiegelman, *ibid.* **60**, 713 (1968).
8. R. Levisohn and S. Spiegelman, *ibid.*, p. 866.
9. ———, *ibid.* **63**, 805 (1969).
10. R. Saffhill, H. Schneider-Bernloehr, L. E. Orgel, S. Spiegelman, *J. Mol. Biol.* **51**, 531 (1970).
11. S. Spiegelman, *Quart. Rev. Biophys.* **4**, 213 (1971).
12. D. R. Mills, N. R. Pace, S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.A.* **56**, 1778 (1966); N. R. Pace, D. H. L. Bishop, S. Spiegelman, *ibid.* **58**, 711 (1967).
13. N. R. Pace, D. H. L. Bishop, S. Spiegelman, *ibid.* **59**, 139 (1968).
14. S. Spiegelman, N. R. Pace, D. R. Mills, R. Levisohn, T. S. Eikhom, M. M. Taylor, R. L. Peterson, D. H. L. Bishop, *Cold Spring Harbor Symp. Quant. Biol.* **33**, 101 (1968); C. Weissmann, G. Feix, H. Slor, *ibid.*, p. 83.
15. D. L. Kacian, D. R. Mills, F. R. Kramer, S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.A.* **69**, 3038 (1972).
16. D. R. Mills, F. R. Kramer, S. Spiegelman, in preparation. The extensive data are being processed for publication.
17. R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, A. Zamir, *Science* **147**, 1462 (1965).
18. F. Sanger, G. G. Brownlee, B. G. Barrell, *J. Mol. Biol.* **13**, 373 (1965).
19. J. Josse, A. D. Kaiser, A. Kornberg, *J. Biol. Chem.* **236**, 864 (1961).
20. D. H. L. Bishop, D. R. Mills, S. Spiegelman, *Biochemistry* **7**, 3744 (1968).
21. The sequence of an oligonucleotide is represented in the 5′ to 3′ direction as an alphabetic string composed of the letters C, A, G, U, which are the 3′-monophosphates of cytidine, adenosine, guanosine, and uridine. Sometimes the location of each phosphate group is indicated by a p to the right of each riboside. Neighbor nucleotides are separated from the oligonucleotide string by a "/". Py is a pyrimidine riboside monophosphate, and pppGp is a 5′-guanosine tetraphosphate. CTP, ATP, GTP, UTP are the corresponding riboside 5′-triphosphates.
22. M. A. Billeter, J. E. Dahlberg, H. M. Goodman, J. Hindley, C. Weissmann, *Nature* **224**, 1083 (1969).
23. R. V. Tomlinson and G. M. Tener, *Biochemistry* **2**, 697 (1963).
24. D. H. L. Bishop, J. R. Claybrook, S. Spiegelman, *J. Mol. Biol.* **26**, 373 (1967).
25. I. Tinoco, O. C. Uhlenbeck, M. D. Levine, *Nature* **230**, 362 (1971); C. Delisi and D. M. Crothers, *Proc. Nat. Acad. Sci. U.S.A.* **68**, 2682 (1971).
26. T. R. Fink and D. M. Crothers, *J. Mol. Biol.* **66**, 1 (1972).
27. O. C. Uhlenbeck, F. H. Martin, P. Doty, *ibid.* **57**, 217 (1971).
28. F. R. Kramer, D. R. Mills, T. Nishihara, P. Cole, S. Spiegelman, in preparation.

29. We acknowledge the assistance of E. Hajjar and J. Mack in preparing the illustrations. Supported by NSF grant GB-17251X2, NIH research grant CA-02332, and American Cancer Society postdoctoral fellowship PF-538.

COVER

First complete nucleotide sequence of a molecule capable of extracellular replication. Each complementary strand is 218 nucleotides in length. In the test tube, one molecule will autocatalytically generate one trillion (1×10^{12}) copies in 20 minutes. See page 916. **[D. R. Mills *et al.*, College of Physicians and Surgeons, Columbia University, New York City]**